DOCUMENT RESUME

ED 359 231 TM 020 000

AUTHOR Elliott, Ronald S.; Barcikowski, Robert S.
TITLE Multivariate Test Statistics and Their Approximations: Some Problems.
PUB DATE Apr 93
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Variance; Computer Simulation; Estimation (Mathematics); Monte Carlo Methods; *Multivariate Analysis; Probability; *Robustness (Statistics); *Sample Size
IDENTIFIERS Biomedical Computer Programs; Hotellings t; Likelihood Ratio Tests; Pillais Trace; Roys Largest Root; Statistical Analysis System; Statistical Package for the Social Sciences; *Type I Errors

ABSTRACT

This Monte Carlo study examines whether, given various numbers of variables, treatments, and sample sizes, in a one-way multivariate analysis of variance, Type I error rates of the test approximations provided by the BMDP program, the Statistical Analysis System (SAS), and the Statistical Package for the Social Sciences (SPSS) for Roy's largest root, Hotelling's trace, Wilks' likelihood ratio, and P. C. S. Pillai's trace meet the stringent criterion of J. V. Bradley (1978) for robustness. For each of 85 conditions, the Type I error rate for each of the above statistical tests was estimated based on 100,000 random samples per situation. Results indicate that in multivariate analysis of variance studies with relatively small numbers of subjects of around 15 per treatment level, or fewer, the current probability levels reported by SAS and SPSS are conservative for the F approximations based on Pillai's trace and liberal for the F approximations based on Hotelling's trace. The BMDP 4V program does not report Pillai's trace and reports accurate probability values for Hotelling's trace. All of these programs report accurate values for the F approximations based on Wilks' likelihood ratio and for Roy's largest root (SPSS does not report Roy's largest root). Recommendations are made for specific conditions. Two tables and five figures present analysis results. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED359231

Multivariate Test Statistics And Their Approximations: Some Problems

Ronald S. Elliott
and
Robert S. Barcikowski

Ohio University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

RONALD S. ELLIOTT

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A paper presented before the American Educational Research Association, Atlanta, April, 1993.

TM020000

Multivariate Test Statistics and Their Approximations: Some Problems

Objectives

The purpose of this of this paper is provide educational researchers with information concerning the Type I error rates of the test statistic (F or $\chi 2$) approximations of the four multivariate statistical tests reported by most statistical packages. The question answered in this paper is:

Given various numbers of variables, numbers of treatments, and sample sizes in a one-way multivariate analysis of variance, do the Type I error rates of the test statistic approximations provided by BMDP, SAS, and SPSS for Roy's largest root, Hotelling's trace, Wilks's likelihood ratio, and Pillai's trace meet Bradley's (1978) stringent criterion for robustness?

Perspectives

Recently we happened upon a data set where using SPSS's program MANOVA (or SAS's program GLM) we found a significant omnibus multivariate F ($p < .044$) based on Hotelling's trace ($\alpha = .05$), but a nonsignificant ($p < .0654$) result for this same statistic when using BMDP's 4V program. Our results are illustrated in Table 1 for SPSS (Hotellings line) and in Table 2 for BMDP (CHISQ line). The difference between the two results was caused by the different approximation methods used by the programs to arrive at their probability values. This led us to this study and to the use of Bradley's (1978, p. 146) criterion for examining the Type I error rates of the approximation methods under a variety of conditions.

Table 1

One-Way MANOVA (Six Treatments, Six Dependent Variables, Five Subjects Per Treatment) Output From SPSS (MANOVA)

Multivariate Tests of Significance (S = 5, M = 0, n = 8 1/2)

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig of F
Pillais	1.11157	1.09582	30.00	115.00	.354
Hotellings	2.78697	1.61645	30.00	87.00	.044
Wilks	.19299	1.32274	30.00	78.00	.164
Roys	.69570				

Note. These results agree with those found using SAS (GLM)

Table 2

One-Way MANOVA (Six Treatments, Six Dependent Variables, Five Subjects Per Treatment) Output From BMDP4V

STATISTIC	F	DF		P
LRATIO = 0.192990	1.32	30.00	78.00	0.1638
TRACE = 2.78697				
TZSQ = 64.1004				
CHISQ = 19.43			30.226	0.0654
MXROOT = 0.695699				0.0197

Bradley's stringent criterion for robustness is that the actual level of significance be within $\pm .1\alpha$ of the nominal level of significance (α). For example, if $\alpha = .05$, then the approximation method should yield a nominal value that falls between .045 and .055. We felt that this level of precision would be necessary for an approximation method found in a statistical package which is meant to be used in a wide variety of research situations.

The *F* approximation for Roy's largest root is based on Harris (1975) in BMDP and Pillai (1965) for SPSS and SAS; for Hotelling's trace, BMDP's chi-square approximation is based on Tiku (1971) and SAS and SPSS's *F* approximation is based on Pillai (1960); Wilks's likelihood ratio *F* approximation for all programs is based on Rao (1973); Pillai's trace *F* approximation is based on Pillai (1960) for SAS and SPSS.

Olson (1976) urged researchers to provide the preceding information in their research reports when he indicated:

> In view of the differing robustness performances of the test criteria, researchers who use the expression *multivariate F* should include a footnote specifying *whose* approximation to *which* criterion was employed. (p. 584)

Methods/Data Source

A Monte Carlo study was undertaken based on a format developed by Olson (1975). A FORTRAN program was written to generate discriminant score data based on no violations of the multivariate analysis of variance (MANOVA) test assumptions. The program used the IMSL (1987) random number generator, RNMVN, to generate random score vectors from a multivariate normal distribution. Given the following finite numbers of dependent variables, treatment levels, and samples sizes, and our investigation of only the .05 level of significance, the preceding method of generating data yields results that are generalizable to any real life situation where the data meet the MANOVA assumptions and where there are no differences among the treatment population means.

The study included the following combinations of dependent variables by treatment levels 2X3, 3X3, 3X6, 6X3, and 6X6. For each combination of dependent variables by treatment levels the equal n sample sizes per treatment level were 4 (1) 20. In each of these eighty-five different conditions the Type I error rates for Roy's largest root based on Harris (1975) and Pillai (1965); Hotelling's trace based on Tiku (1971) and Pillai (1960); Wilks's likelihood ratio based on Rao (1973); and Pillai's trace based on Pillai (1960) were estimated based on 100,000 random samples per situation. Following the logic provided by Robey and Barcikowski (1992), this number of samples would yield power of more than .90 of detecting ($\alpha = .05$) a departure of $.1\alpha$ from a nominal alpha of .05. Our FORTRAN program was run on Ohio's Cray Y-MP Supercomputer.

Results

The Type I error rates based on the *F* approximations found for Roy's largest root and for Wilks's likelihood ratio as well as the $\chi 2$ approximation for Hotelling's trace met Bradley's stringent criterion in all eighty-five cases. However, the *F* approximations found for Hotelling's trace and for Pillai's trace failed to meet Bradley's stringent criterion under various conditions with cell sample sizes between six and fifteen. The latter points are illustrated in figures 1 through 6.

In figures 1 through 5 the actual levels of significance (labeled *probability*) and the treatment sample sizes (labeled *group size*) are plotted based on the *F* approximations for Roy's root, Pillai's trace, Wilks's lambda, the Hotelling-Lawley trace and Tiku's chi-square approximation for the Hotelling-Lawley trace. In these figures Bradley's stringent criterion for robustness is identified by lines at probability values of .045 and .055, and the nominal level of significance is identified by a line at .05. The *F* approximations for Roy's root, Wilks's lambda, and Tiku's chi-square approximation for the Hotelling-Lawley trace all yield estimates of the nominal levels of significance that fall between the lines at .045 and .055, i.e., they all yield estimates of the nominal level of significance that meet Bradley's criterion. However, all five figures also contain treatment sample sizes where estimates of the nominal levels of significance provided by the *F* approximations of Pillai's trace are below .045 and the estimates of the nominal levels of significance of the *F* approximation for the Hotelling-Lawley trace are above .055, i.e., fail to meet Bradley's criterion.

In figure 6 we indicate the conditions and the treatment sample sizes prior to the sample size where the estimates of the nominal levels of significance based on the *F* approximations for the Hotelling-Lawley trace and of Pillai's trace meet Bradley's criterion. For example, given two dependent variables and three treatment levels, the *F* approximation for the Hotelling-Lawley trace yielded an estimate of the level of significance that fails to meet Bradley's criterion when the sample size is 6 or less (see cell 1,1 in figure 6), but would yield an estimate of significance that does meet Bradley's criterion when the sample size is 7 or larger. For the same conditions the *F* approximation for Pillai's trace yielded an estimate of the level of significance that fails to meet Bradley's criterion when the sample size is 10 or less (see cell 1,1 in figure 6), but would yield an estimate of significance that does meet Bradley's criterion when the

Multivariate Test Statistics 6

Figure 1. TWO VARIABLES, THREE GROUPS

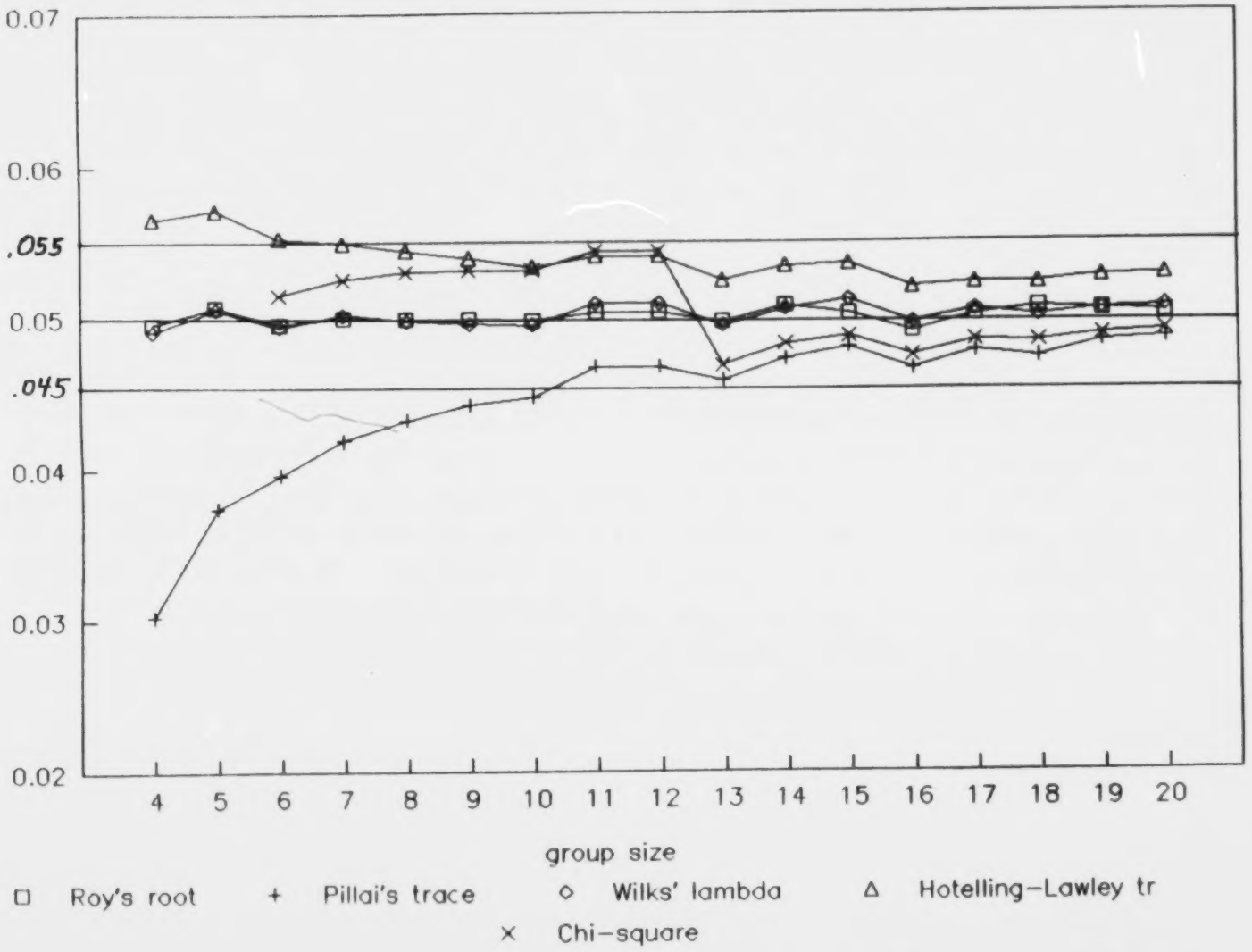


Figure 2. THREE VARIABLES, THREE GROUPS

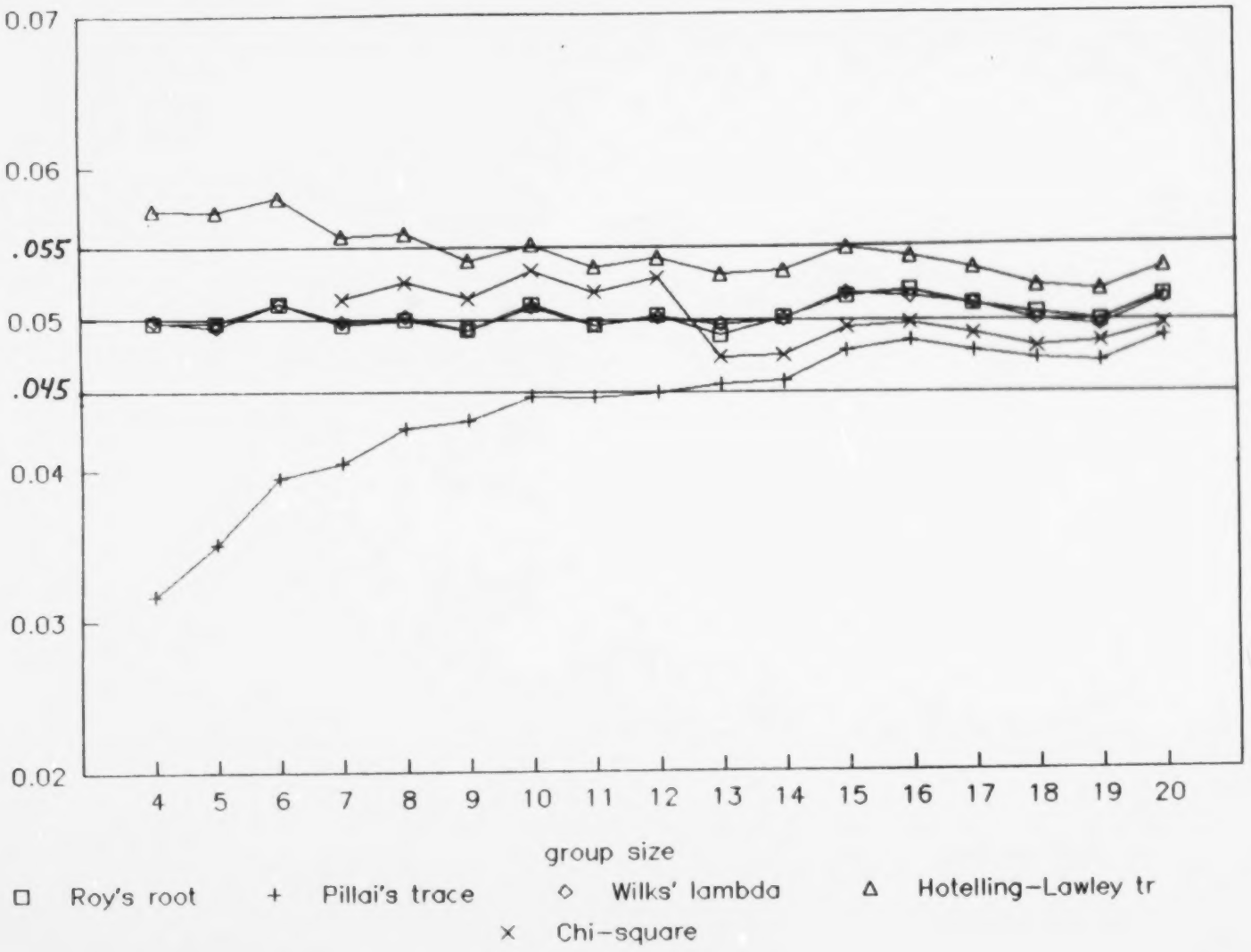


Figure 3. THREE VARIABLES, SIX GROUPS

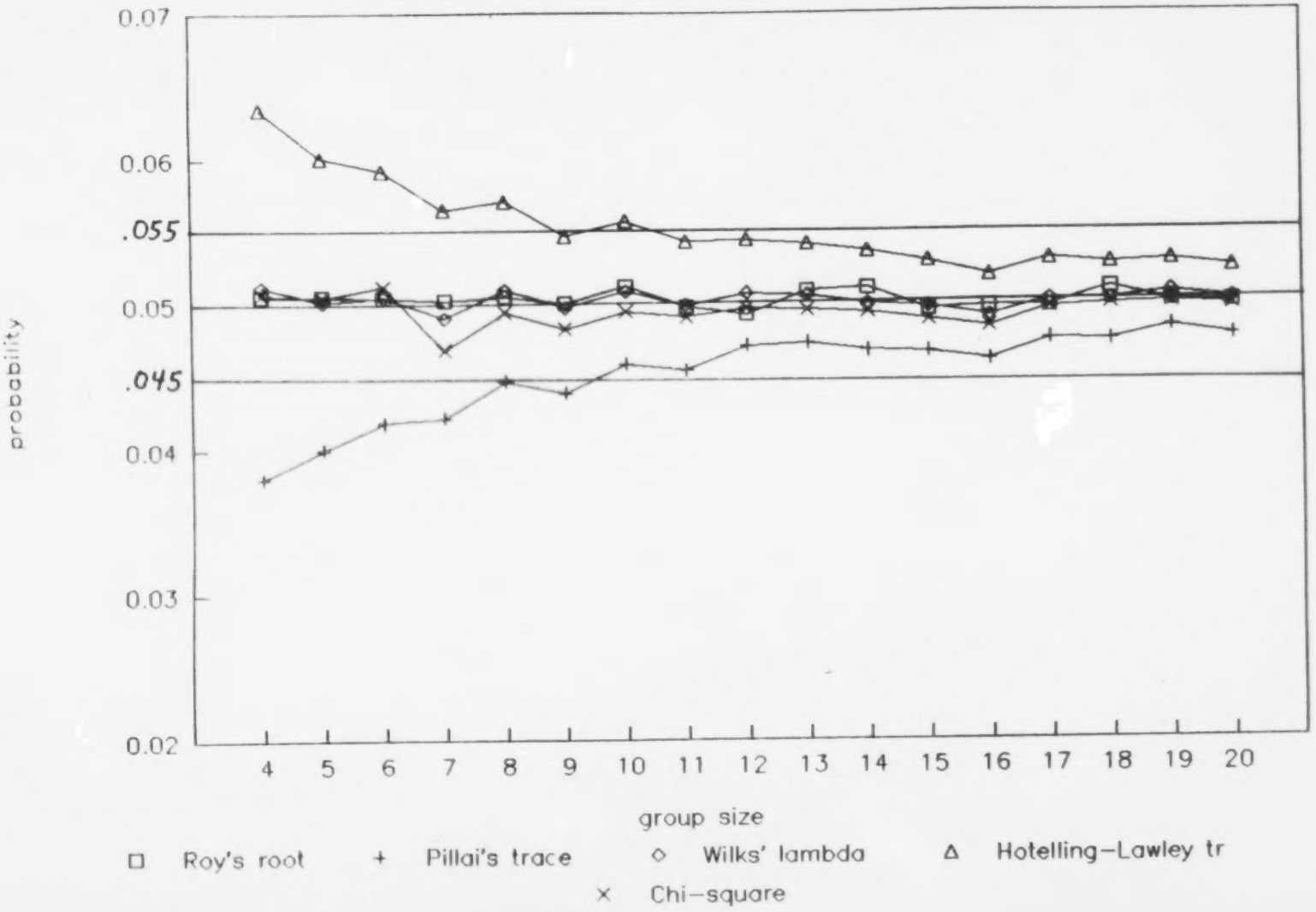


Figure 4. SIX VARIABLES, THREE GROUPS

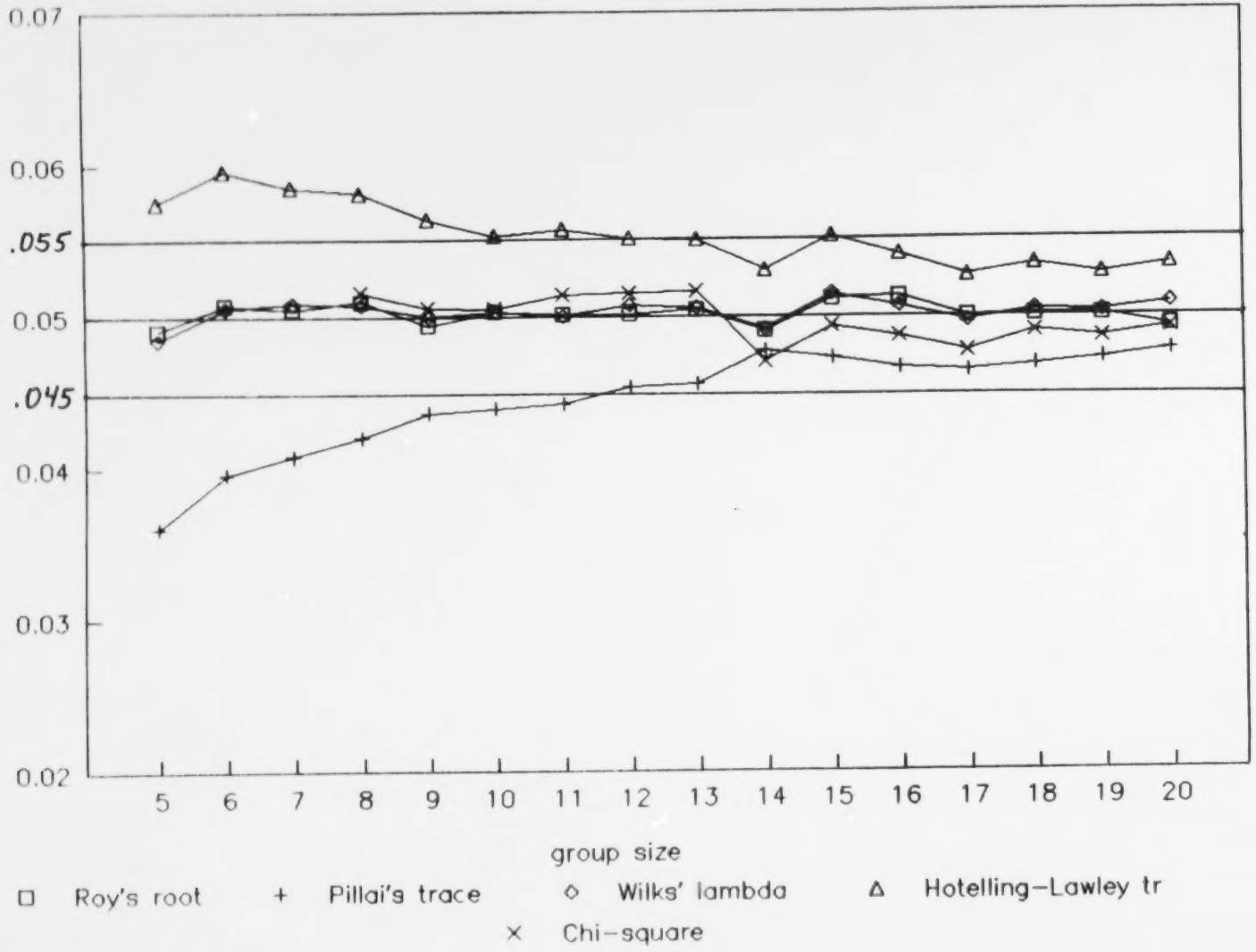
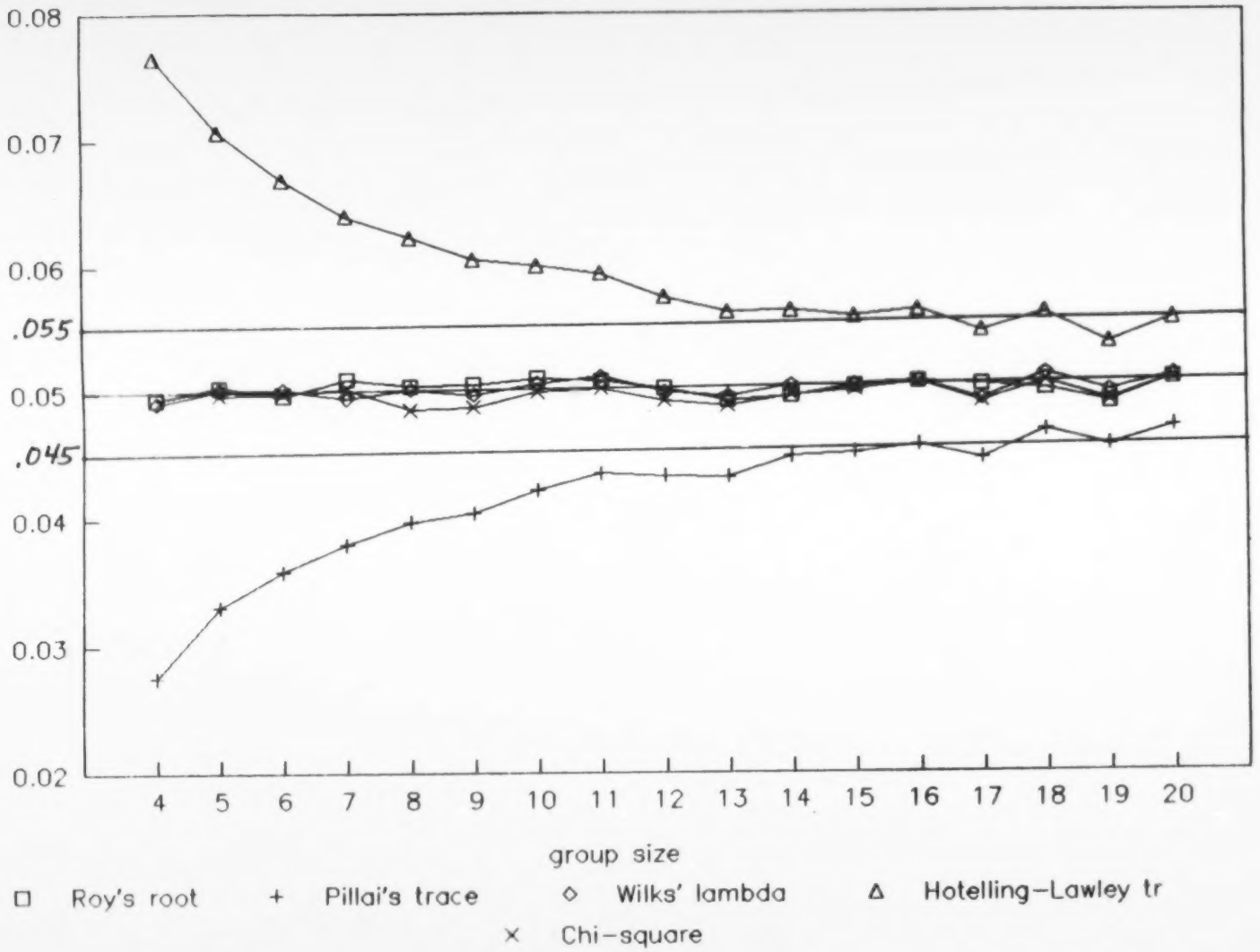


Figure 5. SIX VARIABLES, SIX GROUPS



sample size is 11 or larger. The latter two results can also be observed by viewing the plots for Hotelling's trace and Pillai's trace in figure 1.

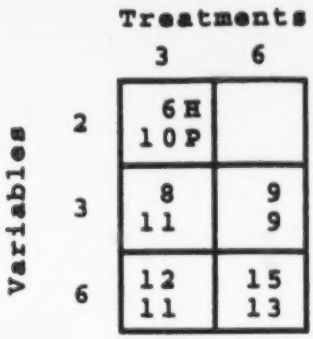


Figure 6. Sample size prior to when the test statistic yielded estimates of the nominal level of significance that meet Bradley's stringent criterion. The first sample size in each cell is based on Hotelling's Trace (H) and the second sample size is based on Pillai's Trace (P).

Conclusions

The results indicate that in multivariate analysis of variance studies with relatively small numbers of subjects of around 15 per treatment level or less the current probability values reported by SAS(GLM) and SPSS(MANOVA) are conservative for the *F* approximations based on Pillai's trace and liberal for the *F* approximations based Hotelling's trace. The BMDP4V program does not report Pillai's trace and reports accurate probability values for Hotelling's trace. All of these programs report accurate values for the *F* approximations based on Wilks's likelihood ratio and for Roy's largest root (except SPSS which does not report a probability value for Roy's largest root). Unfortunately, many MANOVA studies with small subject-variable ratios exist in the social science literature (Olson, 1976). It would be unfortunate if the authors of these studies used Pillai's trace and found no significant result when one existed or used Hotelling's trace and found a significant result when none existed.

Future research. We recommend that future research in this area complete the design shown in figure 6 with the addition of 9 and 12 treatment levels. Furthermore, power

calculations for a wide variety of designs have been recently presented by Muller, LaVange, Ramey, and Ramey (1992) using the *F* approximations for Hotelling's trace and Pillai's trace that we have found yield poor approximations with small sample sizes. Future research which exams the performance of these approximations with large effect sizes and small sample sizes is warranted.

Recommendation. Given a small subject-variable ratio and no violations of assumptions, we recommend that researchers consider the probability values provided by the *F* approximations for Roy's largest root, Wilks's likelihood ratio, or Tiku's $\chi 2$ approximation for Hotelling's trace (as found in BMDP4V). Furthermore, we recommend that all statistical packages use Tiku's $\chi 2$ approximation for Hotelling's trace. Given violations of the MANOVA tests' assumptions (Olson, 1975, 1976), we recommend the use of *F* approximations based on Wilks's lambda (Stevens, 1979) or Pillai's trace (recognizing that this will be conservative). For all conditions, we call for an improved estimate of the *F* approximation for Pillai's trace.

References

Bradley, J. V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology, 31,* 144-152.

Dixon, W. J. (Ed.). (1983). *BMDP statistical software manual, 1985 reprinting.* Los Angeles: University of California Press.

Harris, R. J. (1985). *A primer of multivariate statistics* (2nd ed.). Orlando, FL: Academic Press.

IMSL (1987). *Fortran subroutines for statistical applications.* Houston, TX: IMSL, Inc.

Muller, K. E., LaVange, L. M., Ramey, S. L., & Ramey, C. T. (1992). Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association, 87,* 1209-1226.

Olson, C. L. (1975). A Monte Carlo investigation of the robustness of multivariate analysis of variance (Doctoral dissertation, University of Toronto, 1973). Dissertation Abstracts International, *35,* 6106B.

Olson, C. L. (1976). On choosing a test statistic in multivariate analysis. *Psychological Bulletin, 83,* 579-586.

Pillai, K. C. S. (1960). *Statistical tables for tests of multivariate hypotheses.* Manila: Statistical Center, University of the Philippines.

Pillai, K. C. S. (1965). On the distribution of the largest characteristic root of a matrix in multivariate analysis. *Biometrika, 52,* 405-414.

Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: Wiley.

Robey, R. R. & Barcikowski, R. S. (1992). Type I error and the number of iterations in Monte Carlo studies of robustness. *British Journal of Mathematical and Statistical Psychology, 45,* 283-288.

SAS Institute Inc. (1989). *SAS/STAT user's guide, Version 6, Fourth Edition. Volume 2.* Cary, NC: SAS Institute Inc.

SPSS Inc. (1988). *SPSSX: User's guide* (3rd ed.). New York: McGraw-Hill.

Stevens, J. P. (1979). Comment on choosing a test statistic in multivariate analysis by C. L. Olsen. *Psychological Bulletin, 86*, 728-737.

Tiku, M. L. (1971). A note on the distribution of Hotelling's generalized T^2. *Biometrika, 58*, 237-241.